SUPPLEMENTARY MATERIAL: FURTHER RESULTS FROM OPEN, SENSITIVITY ANALYSES TO COMPARE METHODS OF ADJUSTING RELATIVE RISK ESTIMATES FOR DIETARY MEASUREMENT ERROR, COMMENTS ON SIGNAL ATTENUATION AND UNMEASURED CONFOUNDERS, AND FURTHER NOTES ON THE REGRESSION CALIBRATION ADJUSTMENT

Introduction

In Table 2 of the main text, estimated attenuation factors are presented for a model with log-transformed protein density, potassium density, and energy intakes as the explanatory variables. Two sets of estimates are shown; those based on recovery biomarker measurements are thought to be unbiased, while those based on 24-hour recall assessments may be biased. Supplementary Table 1 shows corresponding estimates of contamination factors for this model, and for a model with one additional nutrient density. In Table 2 and Supplementary Table 1 there are some apparent differences between the estimates based on recovery biomarkers and those based on 24-hour recalls. However, the impact of these differences and their implications for the choice of method to be used to adjust relative risk estimates for dietary measurement error are not immediately clear. Here we examine the implications of the results shown in those tables, using sensitivity analysis. We also give more explanations regarding signal attenuation and unmeasured confounders, regression calibration adjustment, and the design of appropriate validation studies.

Note that the results shown in all tables in the main text and Supplementary Material relate to models with energy, protein and potassium (and in some cases one

other nutrient). Models with fewer nutrient variables, such as protein and energy alone, were also run with similar results (results not shown here).

Specific Aims

The central questions are whether and, if so, how best to adjust relative risk estimates from a nutritional cohort study employing a food frequency questionnaire (FFQ) as the main dietary assessment tool, where the validation data available include one or more 24-hour recalls as the reference instrument? The relative risk estimation methods that are to be compared are: 1) no adjustment for measurement error; 2) univariate regression calibration adjustment based on a 24-hour recall reference instrument, using the method of Rosner et al. (1); 3) multivariate regression calibration adjustment based on a 24-hour recall reference instrument, using the method of Rosner et al. (1).

Regression Calibration Adjustment by the Rosner et al. Method

Suppose that the disease model of interest is as follows:

$$\text{logit}\,(P(D=1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \sum_{k=1}^{p} \gamma_k Z_k \qquad [1]$$

where *D* is the disease indicator, X_1 is log(protein density), X_2 is log(potassium density), X_3 is log(energy), and $Z_1,\ldots,\ Z_p$ are *p* confounder variables that are exactly measured.

In practice the exact nutritional intakes, X_1, X_2, and X_3 are unavailable and are assessed using a FFQ to give values W_1, W_2, and W_3, respectively. Using these in place of their unknown exact values, the logistic regression model [1] is run and estimates

of β_1, β_2, and β_3 are obtained. Denote these estimates by $\hat{\beta}_1^*$, $\hat{\beta}_2^*$, and $\hat{\beta}_3^*$, respectively, where the superscript * denotes that the model is fit to error-prone covariates. We call these the unadjusted estimates of the log relative risks, and they are known to be biased.

Our proposed "univariate" adjustment is to divide each unadjusted estimate by the attenuation factor for that variable, as estimated from the validation study. This is the same as applying the Rosner et al. adjustment for a model with a single error-prone variable to each of the error-prone variables in the multivariate model [1]. The validation study provides, in a subset of participants, reference measurements for X_1, X_2, and X_3, which we denote by R_1, R_2, and R_3, respectively. We assume here that these are 24-hour recall assessments. The estimate for the attenuation factor for the variable W_1, for example, is the estimated coefficient of W_1 in the regression of R_1 on W_1 and Z_1,…., Z_p. Denote the estimated attenuation factors for W_1, W_2 and W_3 by $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\lambda}_3$, respectively. Thus, the three log relative risks adjusted by the univariate method are $\hat{\beta}_1^* / \hat{\lambda}_1$, $\hat{\beta}_2^* / \hat{\lambda}_2$, and $\hat{\beta}_3^* / \hat{\lambda}_3$, respectively. Note that this method uses only attenuation factors to adjust the log relative risk estimates. Contamination coefficients are ignored, assuming them to be close to zero.

Multivariate adjustment by the Rosner et al. method employs estimates of both attenuation and contamination factors, as estimated from the validation study. Denote the vector of unadjusted log relative risk estimates $(\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_3^*)'$ by $\hat{\beta}^*$. We apply the inverse of the estimated "attenuation–contamination" matrix to this vector to obtain the adjusted log relative risk estimates. In our case, this matrix has three rows and three columns. The first row contains the estimated coefficients of the variables W_1, W_2, and W_3 in the

regression of R_1 on W_1, W_2, W_3, and $Z_1, \ldots, Z_p$. The second row contains the estimated coefficients of the variables W_1, W_2, and W_3 in the regression of R_2 on W_1, W_2, W_3, and $Z_1, \ldots, Z_p$, and similarly the third row. Denote this matrix by $\hat{\Lambda}$. The diagonal elements of the matrix are estimated attenuation factors, and the off-diagonal elements are estimated contamination factors. The multivariate-adjusted estimates of the log relative risks are given by the vector $\left(\hat{\Lambda}'\right)^{-1}\hat{\beta}^*$.

Sensitivity Analysis

Model with protein density, potassium density, and energy. The univariate attenuation factors, λ_1, λ_2, and λ_3, estimated from OPEN data were as follows. Using recovery biomarker data as the reference: for men (0.40, 0.49, 0.08); for women (0.32, 0.57, 0.04). Using 24-hour recall data as the reference: for men (0.41, 0.51, 0.23); for women (0.50, 0.58, 0.13).

For the multivariate factors, the diagonal and off-diagonal elements of the matrix Λ are given by the entries in Table 2 of the main text and in the rows for the first three nutrients (energy, protein and potassium) in Supplementary Table 1, respectively. Recovery biomarker-based estimates and 24HR-based estimates are both provided in these tables. Note that although some of the contamination factors are not statistically significant (as noted in the main text), we nevertheless build the simulation on the basis of these non–statistically significant estimates. Due to its limited sample size, the OPEN study did not have enough power to find very small contamination factors statistically significantly different from zero. Therefore, setting all non–statistically significant values

equal to zero would have yielded an over-optimistic evaluation of the performance of the 24-hour recall.

To assess the impact on estimated log relative risks from using adjustment methods based on a 24-hour recall reference, we needed to postulate the true values of β_1, β_2, and β_3. Because these values are, of course, unknown, and because the OPEN study was a stand-alone validation study and not linked to a specific cohort, we chose a series of combinations of β_1, β_2, and β_3, and for each combination we calculated the expected bias that would accrue from use of the 24-hour recall as the reference. Specifically, our steps were: First, choose $\beta = (\beta_1, \beta_2, \beta_3)'$. Second, calculate $\Lambda'\beta$, where Λ is the estimate derived from the recovery biomarkers; this calculation gives the expected value of $\hat{\beta}^* = (\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_3^*)'$ the unadjusted estimates of the log relative risks. Third, apply the three adjustment methods to these values, which yields: $\Lambda'\beta$ for no adjustment; $\left\{(\Lambda'\beta)_1 / \lambda_1^{(24HR)}, (\Lambda'\beta)_2 / \lambda_2^{(24HR)}, (\Lambda'\beta)_3 / \lambda_3^{(24HR)}\right\}$ for the univariate adjustment (ie, each element of $\Lambda'\beta$ divided by the corresponding 24-hour recall–based estimate of the univariate attenuation factor; and $\left(\Lambda_{24HR}^{'}\right)^{-1}\Lambda'\beta$ for the multivariate adjustment, where Λ_{24HR} is the 24-hour recall–based estimate of the attenuation-contamination matrix. Fourth, calculate the difference between each of the values calculated in third step and the true log relative risk, as chosen in the first step; this gives the bias in estimating each log relative risk from each method.

The combinations of values of $\beta = (\beta_1, \beta_2, \beta_3)'$ chosen were as follows. Three relative risk values between an individual at the 90th percentile of intake versus one at the 10th percentile of intake (0.5, 1, and 2) were considered for protein density and potassium

density and two relative risk values (1 and 2) for energy. This gave 18 (ie, 3×3×2) combinations. However, the null combination, where all three relative risks were equal to 1.0, was omitted, leaving 17 combinations in total. The β values for each relative risk were computed by the formula $\beta = \ln(RR)/(2.56\sigma_T)$, where *RR* is the specified relative risk and σ_T is the standard deviation of the true intake on the log scale. The latter was estimated for each of the intakes from data in the OPEN study.

Results were summarized over the 17 combinations for each method, by taking the square root of the mean of the squared biases. We call this "the root mean square bias." We provided this measure also for the restricted set of estimates of coefficients that were truly nonzero (2×3×2 = 12 cases per nutrient), these being the most important from a public health perspective.

Supplementary Table 2 shows the results for the full set of combinations. Both adjustment methods perform better, on average, than the unadjusted method. In addition, the univariate method appears to perform better than the multivariate method. Supplementary Table 3 shows the results restricted to the set of nonzero coefficients. Again, the adjustment methods perform better on average than the unadjusted method, with the improvement larger in this set of cases. For males, the univariate method performs better than the multivariate method, whereas for females, the two methods perform similarly.

Model with protein density, potassium density, carbohydrate density, and energy. To extend the results of the previous section, we investigated biases in log relative risks for a model with four nutrient variables, that is, carbohydrate density in

addition to those in the previous three-nutrient model. With inclusion of this extra variable, the attenuation–contamination matrices increase in size from 3×3 to 4×4.

Because there is no known recovery biomarker for carbohydrates, the true value of its attenuation factor is unknown and, likewise, the true contamination factors for the carbohydrate column of the attenuation–contamination matrix are unknown. We assumed that the attenuation factor is 0.50 and the contamination factors are zero. The other elements were estimated from the OPEN study data. The full attenuation–contamination matrix is shown in Supplementary Table 4, together with that estimated from the 24-hour recall data.

As with the three-nutrient model, combinations of values of $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$ were chosen. Relative risks of 0.5, 1, and 2 between an individual at the 90th percentile of intake versus one at the 10th percentile of intake were considered for protein density, potassium density, and carbohydrate density, and relative risks of 1 and 2 were considered for energy. This gave 54 (ie, 3×3×3×2) combinations. However, we omitted the null combination, where all three relative risks were equal to 1.0, which left 53 combinations in total.

Results were summarized over these 53 combinations for each method, as previously described, and also over the set of nonzero coefficients (2×3×3×2 = 36 cases per nutrient), and are shown in Supplementary Tables 5 and 6. In Supplementary Table 5, for males, the root mean square bias, averaged over all nutrients, was lower for the adjustment methods than for the unadjusted method**,** although for potassium density, the root mean square bias for the multivariate adjustment was not lower than for the unadjusted method. For females, the root mean square bias averaged over all nutrients

was again lower for the univariate adjustment than for the unadjusted method, but this was not true for the multivariate adjusted method. For potassium density, neither adjustment method had a lower root mean square bias than for no adjustment. For nonzero coefficients (Supplementary Table 6), both adjustment methods had, on average lower root mean square bias than the unadjusted method, although once again this did not apply in the specific case of potassium density for females.

Summary and Conclusions Regarding Adjustment of Relative Risks for Measurement Error

Taking all of the results into account while giving primacy to the results in Supplementary Tables 2 and 3, because all the attenuation and contamination factors in those tables were based on real data, we conclude that adjusting the estimates, even with an imperfect reference instrument, such as a 24-hour recall or food record, is a strategy to be preferred to no adjustment. The adoption of the strategy should lead, on average, to better estimation of relative risks in FFQ cohort studies.

In the sensitivity analyses presented, the univariate method of adjustment performed better on average than the multivariate method for males, and the two methods performed similarly for females. However, we have not recommended that investigators always use the univariate method in preference to the multivariate method, for the following reason. The model that we were able to investigate fully, that including protein density, potassium density, and energy, had somewhat larger discrepancies between the biomarker-based and 24-hour recall–based contamination factors than those shown, on average, in Supplementary Table 1. For the 12 factors estimated for this model, the mean

discrepancy was 0.09, compared with a mean discrepancy of 0.06 for the other contamination factors shown in that table. Thus, it is possible that in analyses of other nutrients the univariate method may not show the same advantage over the multivariate method in terms of the root mean square bias seen in the analyses presented here.

This reasoning leads to the recommendation given in the main text, namely: given the current evidence, statistical adjustment of relative risks based on validation data using 24-hour recalls or multiple-day food records as a reference instrument, may be performed using either univariate or multivariate regression calibration. The former method, albeit simpler, should be used only for energy-adjusted food and nutrient intakes, as we did here in the sensitivity analyses.

Although other exposures or confounders of common epidemiological interest, such as hormone use, are also measured with error, we have focused on dietary intake because the general perception is that dietary measurement error is larger, and consequently has greater impact. Physical activity may represent an exposure with similar measurement problems to dietary exposure, and its measurement is now being increasingly researched. Once a body of background knowledge has accumulated with regard to the errors in the measurement of physical activity, it may be possible to provide similar guidelines for the analysis of this exposure and perhaps even the combined analysis of dietary intake and physical activity.

Signal Attenuation and Unmeasured Confounders

We discuss here in more detail the problem of high attenuation, which results in small observed relative risks (eg, 1.15) that are statistically significant because of a large

sample size. As mentioned in the main text, it is difficult to know whether such an association is due to the dietary intake of interest or unmeasured confounders. Deeper analysis of the problem reveals two types of confounder. First, unmeasured confounders in the model that links disease to true usual dietary intake (type A confounders) have the same relative impact on the detection of dietary effects whether or not the dietary intake is measured with error. For example, suppose the true log relative risk (RR) for a specified change in dietary intake is 0.45 (RR = 1.57) and unmeasured confounders cause the log relative risk to be estimated as 0.54 (RR = 1.72) (a 20% overestimation). Suppose also that there is nondifferential dietary measurement error that attenuates the observed log relative risk from 0.45 to 0.15 (RR = 1.16). Then the same measurement error will also attenuate the effect of unmeasured confounding, so that the confounding will cause the 0.15 to be estimated as 0.18 (RR = 1.20), which remains a 20% overestimation. Likewise, if the true dietary effect is null, then although the unmeasured confounders might yield an estimated log relative risk of 0.09 (RR = 1.09) if dietary intake was measured accurately, the same unmeasured confounding would yield a log relative risk of only 0.03 (RR = 1.03) in a model with self-reported dietary intake. This means that unmeasured confounding of this type is not a special concern regarding attenuated dietary relative risks because the effects of the unmeasured confounding are attenuated by the same degree as the dietary effect itself.

Second, unmeasured confounders that are not of type A but are confounders in the model linking disease to reported usual dietary intake (type B confounders) have a potentially stronger effect than those of type A. It seems especially feasible that unmeasured confounding of this sort could give rise to relative risks that, although not

high, are of the same order of magnitude as highly attenuated dietary relative risks. This is because in such cases, the measurement error accounts for a large part of the variation in the dietary report, and this variation arises from unknown sources, some of which could be associated with disease risk. Furthermore, such confounding, unlike that of type A, is not attenuated by the dietary measurement error; indeed, it arises from the measurement error itself.

Clearly, the best way to eliminate such concerns is to find methods of increasing the precision with which we measure dietary intake.

Further Notes on the Regression Calibration Adjustment and the Design of Validation Studies

The simplest form of the regression calibration adjustment involves 1) estimation of the attenuation factor, and 2) dividing the regression coefficient for the dietary intake of interest (derived from running the disease model with reported dietary intake) by the estimated attenuation factor. Because attenuation factors are small (often as low as 0.3), this adjustment can cause an appreciable increase (often as much as threefold or higher) in the estimated coefficient. Consequently, it is important that the attenuation factor is estimated fairly precisely, because a large error, especially one that reduces the estimated attenuation factor, can lead to unreasonably inflated estimates of dietary effects together with very wide confidence intervals. Therefore, validation studies should include at least several hundred participants to enable the attenuation (and contamination) factors to be estimated with acceptable precision, with standard errors less than 0.05 whenever possible.

A second challenge to regression calibration adjustment is related to the type B confounders described in the previous subsection. These confounders create a challenge to the regression calibration adjustment because they induce differential measurement error, which in turn causes bias in the estimated attenuation factor. As mentioned in the main text, differential measurement error is less likely to occur in cohort studies than in case–control studies. However, it cannot be completely ruled out and investigators should be aware that its effects could possibly reduce or increase the estimate of the attenuation factor by 25%, or, at the maximum, up to 50%. For this reason, attenuation factors for energy-adjusted intakes that appear unusually low (say, less than 0.20) should be viewed with caution. In these cases, the regression calibration adjustment will yield very uncertain estimates.

A further challenge to the regression calibration adjustment arises from the difference in the period covered by the main study instrument (FFQ) and the reference instrument. While the FFQ usually covers a period of several months up to 1 year, the reference instrument (usually 24-hour recall or multiple-day food record) covers a period of 1 day up to several days. If the intake of interest is a food or nutrient that is consumed regularly and at approximately the same frequency throughout the year, and the reference instrument is applied within the period covered by the FFQ (or even shortly before or after this period), then one can assume that the intake reported on the reference instrument measures the same level of intake that occurred during the period covered by the FFQ. However, if the intake of interest has seasonal variation, as can occur with fruits or vegetables that are available to the population only at specific periods during the year, then the timing of the reference instrument must be chosen more carefully. In such cases,

the reference instrument should be applied so that there is a uniform distribution of applications across the whole year (ie, approximately equal numbers of participants should complete the reference instrument in each season of the year). Note that it is not necessary that each participant complete the reference instrument in each season. Indeed, a single administration of the reference instrument to each participant is sufficient, as long as these administrations are evenly spread out over the seasons. This precaution will allow unbiased estimation of the attenuation coefficient that is required in the regression calibration adjustment from the validation study data. In addition, to offset the extra within-person variation in the daily consumption of these nutrients, it is advisable either to increase the sample size of the validation study or to include repeat assessments of participants using the reference instrument.

The remarks above regarding the need to accommodate seasonal variation in consumption in the design of the validation study apply generally to other variations in intake over time (eg, the tendency for persons to consume more on weekends than on weekdays). If the intake of interest is a food or nutrient that is consumed episodically (ie, one that is not consumed every day by a substantial proportion of the population), then more care is required with the validation study design. In this case, it is advisable to include repeat measurements of the reference instrument in at least a subgroup of the participants. Moreover, this subgroup should be large enough that a reasonably large number (ie, at least 50 participants) report intake of the target food or nutrient on more than one administration of the reference instrument. However, not every participant needs to report intake of the target on at least one administration of the reference instrument. The special methods that are used to estimate the attenuation coefficient for this type of

intake are described by Midthune et al. (2), and software for conducting such analyses is available online at http://riskfactor.cancer.gov/diet/usualintakes/.

Sometimes linear regression calibration may not provide the best prediction of usual intake and a nonlinear model is more appropriate. In these cases, nonlinear regression calibration may provide increased power to detect dietary–disease associations. This matter is a topic for further research.

REFERENCES

1. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol, 1990; 132(4):734-45.

2. Midthune D, Schatzkin A, Subar AF, et al. Validating a food frequency questionnaire for intake of episodically consumed foods: application to the National Institutes of Health-AARP Diet and Health Study. *Pub Health Nutr*, doi:10.1017/S1368980011000632. Published online by Cambridge University Press 13 April 2011.